Crusoe Processor Benchmark Tech Note

Crusoe™ Processor Systems and ZD Labs Winstone™ 99 and BatteryMark™ 3.0 Benchmarks

Transmeta Corporation

February 18, 2000

Property of:
Transmeta Corporation
3940 Freedom Circle
Santa Clara, CA 95054 USA
(408) 919-3000
http://www.transmeta.com

Overview

The Transmeta Crusoe processor is based on a set of technology innovations [1] that provide x86 software compatibility, high performance, and low power operation for mobile computing platforms. The Crusoe Code Morphing™ software and LongRun™ power management technologies incorporate intelligent adaptive processes that optimize performance and power consumption dynamically during software execution. Because these adaptive processes are monitoring system and software activity, and modifying the Crusoe processor operating characteristics in response to user activities, the Crusoe processor is sensitive to the “realism” of benchmark software activities. “Realism” in this context means the benchmark software must duplicate, to a much greater degree than with previous non-adaptive processor technologies, realistic and characteristic system activity.

This requirement for a much greater degree of benchmark “realism” is necessary if the benchmark is to report results that correlate with user experience. Because current industry standard benchmarks were developed before the introduction of Crusoe’s adaptive processor technologies, these benchmarks did not anticipate the strict Crusoe requirement for much more realistic benchmark operation. Existing industry standard benchmarks will necessarily have to be updated to address the Crusoe requirement for “realistic” system activity, and if not updated, will not be useful benchmarks for Crusoe processor-based system characterization.

ZD Labs benchmark operation has developed a series of popular benchmarks that are widely used in the PC industry. Winstone 99 is an application performance benchmark that executes a suite of standard Windows® PC applications, driven by automated user input scripts. BatteryMark 3.0 is a battery life benchmark for use on mobile Windows PCs. In combination, these two benchmarks are currently the most popular ZD Labs benchmarks for mobile computer evaluations.

The majority of standard PC industry benchmarks are unsuitable for Crusoe processor-based system characterization, for a variety of reasons. Many are synthetic (no real application code) benchmarks that exhibit poor correlation with user perceivable system operating attributes. Application-level PC benchmarks, such as Winstone 99, are also unsuitable, in their current form, for Crusoe processor characterization as described below. Similarly, the popular ZD Labs battery life benchmark, BatteryMark 3.0, has serious shortcomings for Crusoe processor-based mobile platform characterization. The issues with these two popular ZD Labs benchmarks are outlined below, and the reasons they are poor choices for characterizing Crusoe processor mobile platforms is detailed.

BatteryMark 3.0

BatteryMark 3.0 is intended to measure the expected battery life of mobile PCs. This benchmark is a synthetic battery run-down test with a single usage profile. BatteryMark 3.0 uses no real application code, and runs with an arbitrary fixed usage profile, 25% active and 75% idle. This workload and usage profile is uncharacteristic of real application interaction with the power management system of mobile PCs, and is so narrow in scope, that it is highly non-representative of the vast variety of actual mobile platform usage scenarios and real workload requirements. The benchmark is of limited usefulness for system-level mobile Windows PC battery life testing due to it's poor correlation with many users' actual battery life observations.

The fundamental issue with using BatteryMark 3.0 to evaluate Crusoe processor mobile systems is the degree or lack thereof of "characteristic" (realistic) system activity. Because of the complex and adaptive nature of today's mobile PC power management systems, and the adaptive technology of the Crusoe processor LongRun power management, it is essential that benchmark software exercise the system in a realistic manner. It not, the system will not utilize it's power management system in a "characteristic" way, and not use the system power management features in a way reflective of real usage. Characteristic operation means operation representative of normal system operation under normal user operating conditions.

Benchmarks that do not utilize the system in characteristic operation cannot be expected to generate results that reflect real-use system behavior, and are thus poor choices for system benchmarking. Adaptive power management technologies place stricter demands on benchmarking methodologies to test the adaptive systems in a manner characteristic of real use operation. BatteryMark 3.0 does not exercise the system and LongRun adaptive power management technologies in a realistic way, representative of real mobile application workloads and real-user operation, and is thus not an effective benchmark for Crusoe processor-based mobile systems.

Winstone 99

Winstone 99 is the application performance benchmark most commonly used in the PC industry to evaluate both desktop and mobile PC systems. The standard procedure used in many Ziff-Davis publications' mobile PC testing is to run Winstone 99 on the mobile system plugged into the wall and not operating off battery power. The system power management is also turned off to force the machine into a configuration favoring peak performance, but highly unrepresentative of system operation on battery power with full power management enabled. Since truly mobile platforms are intended to run primarily, if not exclusively, under battery power, this methodology fails to report the expected end-user performance in battery-powered mode.

Winstone 99 may, of course, be run on mobile systems under battery-powered operation with the system power management features fully enabled. This methodology (battery operation/full power management) is a much more realistic indicator of the "untethered" operating characteristics of mobile Windows PCs. For traditional x86-compatible processors, the use of Winstone 99 for mobile system characterization using the battery-powered/power-managed test methodology is sufficient to provide a result with reasonable correlation with user experience. For Crusoe processor-based mobile Windows platforms, however, there is still a major issue with Winstone 99 in this scenario.

The reason that Winstone 99 is unsuitable for characterization of Crusoe processor mobile systems stems from the nature of the adaptive software technologies underlying the Crusoe processor. In the course of a detailed analysis of the operating characteristics and execution profiles of a large range and variety of standard PC software, Transmeta engineers have spent considerable time observing the run-time behavior of a wide variety of PC benchmark programs. The tools and instrumentation developed for analyzing and optimizing the Code Morphing software for the Crusoe processor have also provided a unique window into the operation of PC benchmarks as well as standard application programs.

The results of this analysis for Winstone 99 indicate the methodology used in this benchmark fundamentally distorts the operating characteristics and performance of the software applications used in the benchmark on systems with the Crusoe processor. The applications in the benchmark are being driven by the benchmark scripts in an uncharacteristic accelerated operating mode, such that they operate within the benchmark totally differently than in normal stand-alone user operating mode. This behavior is only apparent on Crusoe processor systems, and is an artifact of the accelerated script-driven test methodology in relation to the Code Morphing dynamic software translation technology. The applications within the Winstone benchmark operate in a much higher performance mode when used in typical user-paced working sessions than they do using the Winstone 99 test scripts. To understand the Winstone 99 benchmark issues, you must know a few details of the operation of the Crusoe processor Code Morphing technology. A short description is offered here; for more details see "The Technology Behind Crusoe™ Processors" white paper [1].

There are two features of Code Morphing software operation that are significant here. The first is that commonly executed application code is translated into very fast native instruction sequences that execute with very high peak performance. The second is that making the translations takes some amount of the processors available operating time. Making translations typically consumes some very small percentage of processor overhead. The payoff in making translations comes when the translated code is executed. The more often a given translation is executed, the greater the payback for the time invested making the translation. This is exactly the area of Crusoe processor operation that Transmeta engineers have spent so much time analyzing and optimizing.

The next significant piece of information is the way users typically, almost without exception, interact with their applications and systems. When users open applications, navigate around within applications, enter data, read data, and all the other things that users do, there are many opportunities

for the processor to find idle time to make translations and consume the required translation overhead in a manner invisible to the user. In addition to the user-paced nature of many PC applications, users typically move through sets of features and functions in a repetitive manner, and spend significant amounts of time within a relatively limited set of software functions. This is the basis for two important performance enhancing features of PCs, the first, which applies to all PCs today, regardless of processor type, is cache behavior. If software did not stay in relatively localized functions and repetitively execute code for significant amounts of time, cache memory would not provide the huge performance speed-up so characteristic of current cache-based processor systems. The second performance enhancing feature, unique to Crusoe processor systems, is translating commonly used code and executing the translations repeatedly at high speeds.

With this background in Code Morphing software translation and typical user application behavior, the methodology flaws in Winstone 99 can be exposed. Winstone 99 contains application scripts that send keystroke and mouse movement information to the applications within the test suite. There is nothing inherently questionable with this particular scripting technique, other than the fact that these scripts run thousands of times faster than humans can possibly interact with the applications. For a conventional desktop processor, with no translation capabilities, the application behavior at this accelerated user interaction rate is the same as when operating at the much slower typical user interaction rate. The benchmark result is thus not skewed relative to the normal operation of the application. A conventional processor derives no benefit from the extra time available between keystrokes.

With the Crusoe processor, however, the accelerated test scripts introduce a distortion in the result because the Code Morphing process is uncharacteristically disrupted by the lack of time available for translation and optimization. In normal system operation, with a human-paced interaction scenario, the Crusoe processor, unlike conventional processors, can use system idle time to do translations and optimize it’s execution performance. Because the accelerated scripting methodology of Winstone 99 does not allow the Crusoe Code Morphing software to utilize the translation and optimization features as it normally does during real user operation, the score reported by Winstone 99 is lower than the user perceived system operating speed. Winstone 99, in its current form, is thus unsuitable for characterization of Crusoe processor-based systems. If the user interaction scripts in Winstone 99 were slowed down to operate at normal user interaction speeds, the measurement flaw of this benchmark with respect to the Crusoe processor can be corrected, and the modified benchmark would then be suitable for Crusoe processor system characterization.

Conclusion

The Crusoe processor was developed to address the needs of mobile computer users, and it should be measured using the attributes important to mobile users. Mobile computer benchmarks should measure performance in combination with the energy consumption penalty for that performance, and

they should address the issue of performance sufficiency as well as peak performance. Mobile benchmarks should measure performance and energy consumption using real mobile application workloads in mobile systems configured and operating the way users configure and operate them, under battery power. In the absence of adequate industry-standard mobile benchmarks today, Transmeta has developed a new benchmark methodology [2] for evaluating mobile computer products. Benchmark results for the Crusoe processor using this new methodology are reported in [3].

Standing aside from the particular issues and limitations of standard PC industry benchmarks such as Winstone 99 and BatteryMark 3.0, it is clear that the new adaptive technologies in Crusoe processors require a careful approach to system characterization. The essential requirement of realistic and characteristic mobile system operation during Crusoe processor-based system benchmarking must be met in order to generate results reflective of actual end-user experience. This is a new requirement for benchmark realism that, prior to the Crusoe processor, was not of such fundamental significance to benchmark accuracy. It is expected that independent benchmark development organizations will be sensitive to the issues described above with respect to Crusoe processor adaptive technologies, and future releases of standard industry benchmarks will measure system performance and energy consumption in realistic and characteristic application usage scenarios.

References

[1] Alexander Klaiber, "The Technology Behind Crusoe™ Processors", *Transmeta Corporation White Paper, January 2000*.

[2] Daniel McKenna, "Mobile Platform Benchmarks", *Transmeta Corporation White Paper, February 3, 2000*.

[3] "Crusoe™ Processor Benchmark Report", *Transmeta Corporation*.

To learn more about mobile platform benchmarks, visit http://www.transmeta.com.